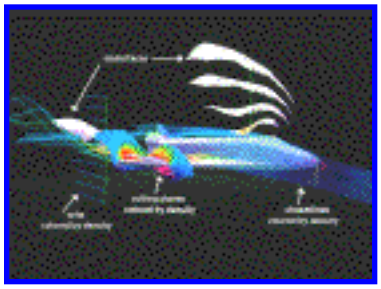
NAS News, Nov-Dec 1995



NAS Will Play Pivotal Role in Agency Supercomputing Consolidation, NCITE

by Walt Brooks,NAS Division Manager

NASA is at a strategic cusp in its evolution. Roles and missions within the agency are being clarified, and one important outcome is that Ames Research Center was recently named the lead center for Information Technology. To help clarify this new role, a plan has been submitted to NASA administrator Dan Goldin to initiate the NASA Center for Information TEchnology (NCITE), whose goal is the development of information technology capabilities necessary to "revolutionize NASA missions in the 21st century." As part of the effort to significantly streamline our operations and to consolidate roles and missions, Ames has also been named the lead center for NASA's scientific and engineering supercomputing.

The NAS mission fits extremely well within Ames' role as the lead center for Information Technology. Within the last two years, the NAS Systems Division has taken on many new responsibilities, including setting up testbeds and undertaking major systems software development efforts for the High Performance Computing and Communications Program, as well as developing and operating the Aeronautics Consolidated Supercomputing Facility. This increase in scope and the resulting significant operational savings represent some initial steps toward realizing the NASA-wide supercomputing consolidation, now underway. (See related article.)

NAS will play a pivotal role in NCITE and other supercomputer efforts for the agency. The division will spearhead efforts in petaflops computing and nanotechnology and will continue to push for balanced systems that will enable us, at the turn of the century, to complete a full multidisciplinary analysis of an entire aircraft. Teraflops and petaflops performance will be necessary to implement desired resolution and accuracy. Although early massively parallel processor architectures failed to live up to their promises, these prototypes helped uncover some underlying technological hurdles, and recent systems have -- for limited applications -- demonstrated sustained performance equal to or greater than the traditional vector symmetric multiprocessor architecture. To reach the teraflops and petaflops goals, an increase in parallel computing scalability will be required, by several orders of magnitude. The NAS Division will be at the heart of efforts to develop tools and testing techniques that will allow us to move forward in this area.

Major challenges ahead include the issues of High Performance Fortran compilers, "automatic" message-passing systems, and effective parallelization. We believe that developments in storage technology and

the evolution of some of our testbed machines will allow us to continue to target aggressive increases in computing speed. As always, today's supercomputer is tomorrow's workstation. With this in mind, NASA has dedicated some computers to prove the feasibility of emerging wind tunnel test techniques, such as phased array acoustic sensors. Feasibility demonstrations in this area have only a few years' wait before these applications will be available on high-end workstations.

NASA is facing some of the same financial challenges that industry has grappled with over the past decade, including a 20 percent reduction in our 1997 budget. The resulting dramatic contraction in supercomputer investments will further fuel the move to cost-effective, commodity-based architectures. These machines will be driven primarily by business applications, such as databases and decision-support systems. NAS and the other key supercomputing centers will play a vital role in influencing the development of these architectures as much as possible and then will adapt the resulting systems to the unique requirements of high-end scientific computing. The goal for the NAS team is to try to leverage this emerging architecture to significantly reduce cost of ownership while pushing next-generation machines to achieve 20-30 gigaflops (sustained) with 1 teraflops (peak) performance.

This is an extremely exciting time. NAS will be challenged to raise its horizons and move super-computing forward for the entire agency. In the near term, we face a new set of commodity-based technology challenges that are critical to technology evolution. Finally, we will use what we have learned from the last two years of consolidation to help keep NASA on the technology frontier, though operating within a significantly smaller budget.

Next Article | Contents | Main Menu | NAS Home

NASA Pursues Consolidation of Supercomputing Resources

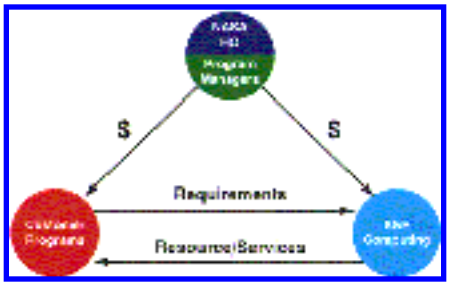
By Eric A. Hibbard

One of the recommendations of NASA's FY94 Zero Base Review was to "Define and consolidate the Agency-wide requirements definition and implementation strategy for the provision and management of large-scale computing services at a single center." In June 1995, John C. Lynn, NASA Chief Information Officer (CIO), asked Ames Research Center to lead the development of a plan for consolidating the management of the agency's supercomputing resources. This supercomputing management plan was forwarded to NASA management by the August 1 deadline, following a week-long meeting of NASA's Ames, Langley, and Lewis Research Centers; Goddard and Marshall Space Flight Centers; Johnson Space Center; Jet Propulsion Laboratory; and NASA headquarters representatives. In late August, Ames was selected as the lead center for supercomputer management and was tasked with developing an implementation plan by December 4, 1995 that would become effective on October 1, 1996.

The NASA Supercomputing Consolidation effort began with an inventory of all the agency's high-performance computing assets, including traditional central computer facilities, the NAS Facility, and the High Performance Computing and Communications Program's parallel and cluster system testbeds. Approximately 40 systems, with a combined estimated purchase price of $300 million, were initially identified as performing supercomputer or related functions (for example, mass storage). Pending further analysis, some of these systems may be removed from consolidated management; they currently deliver an estimated 1.1 million CRAY C90 (1 CPU) equivalent hours annually.

Analysis Reveals Key Problems

Initially, the consolidation was driven by budgetary considerations -- to save an estimated $25 -- $30 million a year in the area of supercomputing for three to five years. However, the findings indicated that the agency's information technology budget cuts, starting in FY97, would already have a negative impact on supercomputing activities. Further analysis concluded that NASA probably could not afford the costs associated with existing supercomputers at its various centers. Consequently, the consolidation effort has gained importance, as the agency must now take steps to ensure viable supercomputing resources while maximizing those funds available for supercomputing.



A key issue identified during this process was that NASA senior management has had limited opportunities to participate in allocating existing supercomputing resources, making it difficult to ensure that the agency's priorities are being addressed. Their primary influence has been through funding, but the requirements for the supercomputing resources and the benefits derived from their usage are fairly hidden (see Figure 1). The net results of this approach include:

- It is difficult to ensure that NASA's priority projects get access to its supercomputing resources.
- Supercomputing resources are not always used to solve NASA's critical problems, but instead, to solve the centers' critical problems. For example, a center with good resources might choose to run a set of problems (that run just as easily on workstations) on a supercomputer to improve turnaround time.
- The return on investment in supercomputing is unclear.
- It is difficult for NASA to buy "smart" since each center negotiates its own purchases.

Consolidation Has Many Benefits

By consolidating the management of supercomputing resources (though not necessarily the systems), the supercomputing component of NASA scientific and engineering (S&E) computing can meet the following goals:

- Be responsive to national aeronautics, Earth and space science, and space flight community priorities.
- Provide an S&E computing capability that balances cost with researcher productivity.
- Within budget constraints, meet customer expectations in all areas, including data management, visualization, and environments to set up the jobs.
- Provide an integrated computing capability supporting technology development, science, management, and business.
- Stay on the cutting edge -- numerical simulation is key to NASA innovation and productivity.
- Use the right computer for the right job.
- Use precious NASA resources (work force, dollars, facilities) to support supercomputing, where

required.

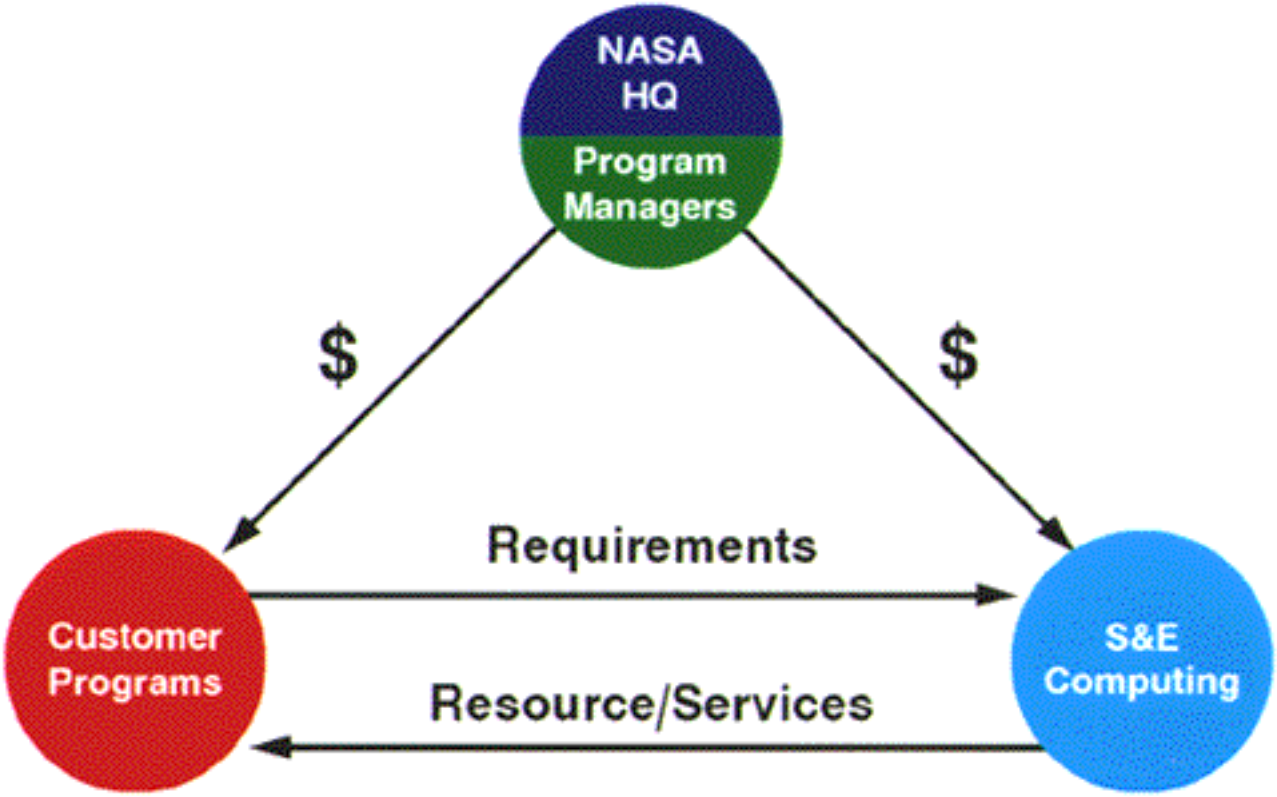
Budget projections indicate that there will be insufficient supercomputing resources to meet the demands of NASA programs in FY97, and senior NASA management will get directly involved in establishing usage priorities. The consolidated management approach will also allow NASA to take advantage of its purchasing power.

User Access Based on `Best Fit'

From a user perspective, an allocation will likely provide access to any one or all of NASA's supercomputer resources -- based on the concept of "best fit for the class of problem." However, the allocation will be granted as part of a project's overall resources -- similar to the way funding is allocated. Many of the details have yet to be worked out, and a significant amount of technical work lies ahead to make the process work seamlessly.

For more information on NASA's supercomputing consolidation effort, send email to Henry Lum, Deputy Director of Information Systems at Ames, at Henry_Lum@qmgate.arc.nasa.gov, or send email to Bruce Blaylock, NAS chief scientist, at blaylock@nas.nasa.gov.

Next Article | Contents | Main Menu | NAS Home



Overview of current NASA supercomputing requirements, funding, and allocation processes.

to the article

NAS Ports Mass Storage to Lower-cost Platforms, Maintains Performance

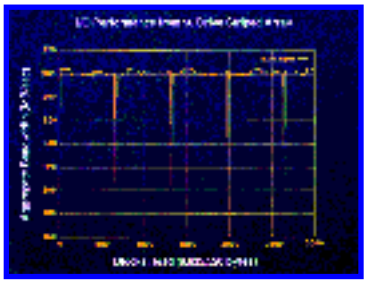
by Elisabeth Wechsler

A project team at the NAS Facility has developed a cheap and effective solution to mass storage, which traditionally has used specialized mainframe computers. "The goal is to port NAStore [NAS's mass storage system] software to less expensive platforms and also to see if we can get the same kind of performance," said Alan Poston, staff scientist and project lead.

"By using fast/wide SCSI II [Small Computer Systems Interface] disk drives to build high-performance disk arrays, you gain the advantage of commodity pricing," he said. A disk array is a collection of disk drives connected by, for example, a number of SCSI buses.

"For about one-tenth of the current price, we get the same storage capacity at 140-150 percent performance improvement. Plus, we could avoid $1 million a year in maintenance costs for NAS's current mass storage system," said Karl Schilke, a student intern working on a part of the project involving the NAS Virtual Windtunnel (VWT).

Performance Exceeds Expectations



Performance levels are higher than expected. The original requirement was 500 megabytes per second (MB/sec) but 250 MB/sec was the performance level agreed to for the NAS VWT testbed. Initial data rates (at 270 MB/sec) exceeded those for NAStore's Convex system (200 MB/sec), and subsequent results have exceeded 300 MB/sec, according to Schilke.

Using Silicon Graphics Inc.'s (SGI) IRIX 6.1 (a version of UNIX) with XFS, it's possible to set up drives in a striped configuration. "Striping is a divide-and-conquer technique," Schilke explained. "By spreading the data across many disks and reading it back in parallel, we can achieve very high aggregate bandwidths."

"The VWT uses the array as a buffer cache, reading CFD datasets at 300 MB/sec, and visualizing it in real time with special virtual reality software," he continued. "The VWT techniques could be used in other applications that create large datasets for visualization in near real time."

"In the past, many datasets were so large that they couldn't be loaded from disk fast enough to be displayed interactively. The VWT overcomes this limitation, allowing scientists to fully explore and understand their datasets," Schilke said.

The testbed, which started in July, used an SGI Power Onyx workstation with 96 drives on 24 SCSI buses. This workstation is similar to the Power Challenge L model used for the SGI cluster testbed at NAS, but has more memory and contains two high-performance 3D graphics engines. The work on this phase was completed in mid-September.

Higher Overall Bandwidth

"Although each drive has a lower bandwidth than the current Convex drives, the overall aggregate bandwidth is as high or higher, and the cost is about 40 percent less," Poston explained, adding that with the Convex mass storage system, NAS "paid a premium to get high-performance disk drives."

The next step is to look for the most cost-effective computer platforms to support these disk arrays, he said. Work on an Intel Pentium personal computer platform was started in February and is expected to be completed at year-end, Poston said. Jason Thorpe, another student intern, is assigned to this phase of the project. An additional effort that involves porting NAStore to the Digital Equipment Corp. (DEC) Alpha was scheduled to begin in October.

NetBSD, a free UNIX-like operating system based on 4.4 BSD-Lite (which runs on Intel and DEC Alpha platforms), provided the basis for a striped disk device, which Thorpe reworked and adapted. Once the modifications to the NetBSD kernel are complete, any platform that NetBSD supports (currently 12) will be able to run NAStore.

Thorpe said that some of the technical issues addressed by the use of NetBSD in the NAS mass storage testbed have included reliability, performance, and access to operating system source code.

UNIX-like Operating Systems

Up to now, commodity hardware has not offered adequate performance levels, Thorpe said. High-performance CPUs such as the Pentium and Alpha are changing that with the PCI [Peripheral Component Interconnect] bus, a high-bandwidth bus for personal computer architectures, he added.

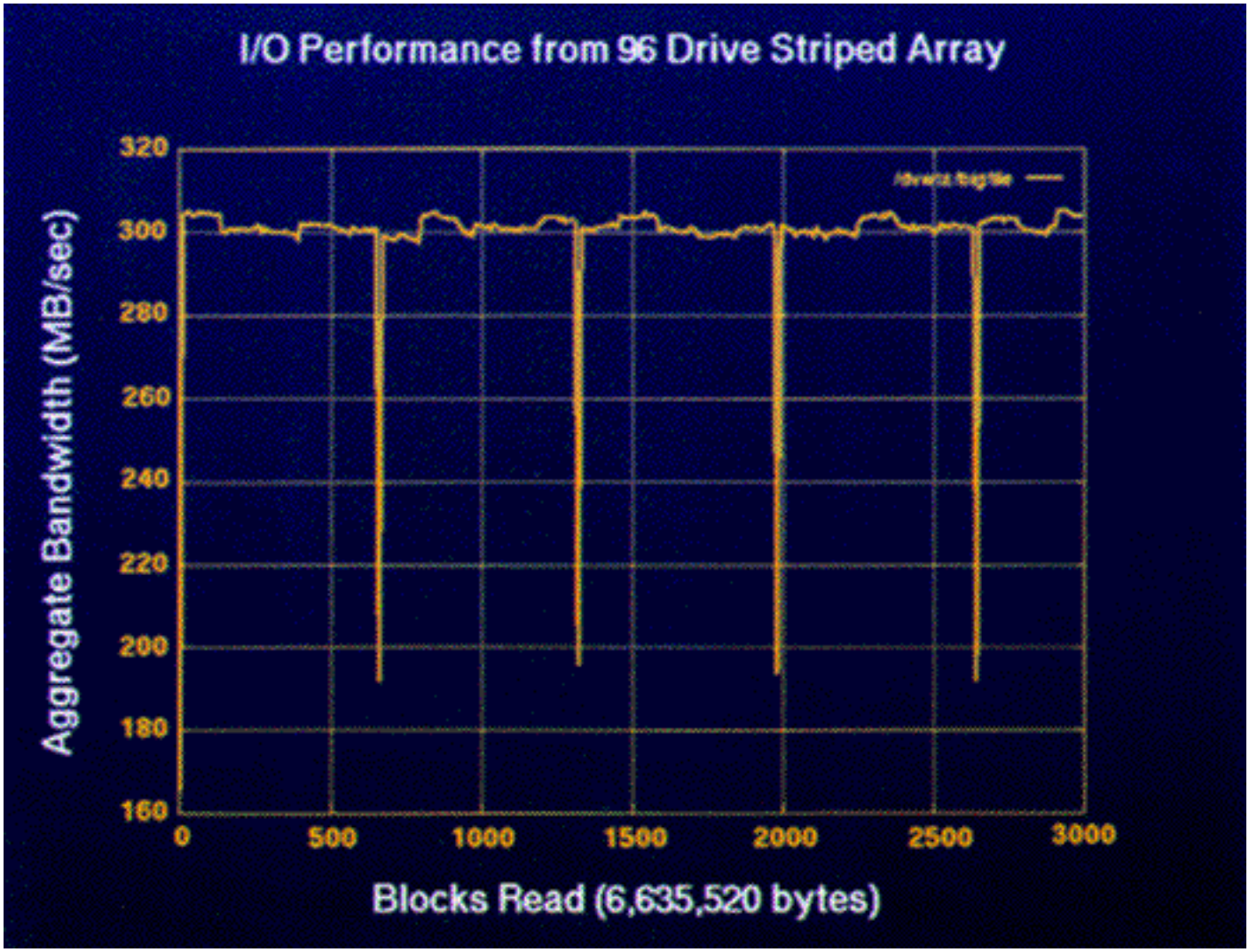
The advent of free UNIX-like operating systems has also helped with using commodity hardware. "Before, you could use Berkeley UNIX on a VAX or System V, but the source code was not generally available," he said.

The impact of these projects at NAS will be that, instead of one big machine to control mass storage, "we'll eventually see a variety of distributed mass storage on smaller machines, better cost-per-megabyte

investment through commodity hardware, and housing more storage capacity with a smaller footprint," Thorpe said.

The team hopes to make the new mass storage system available for widespread use by September '96. For more information, send email to Schilke at rat@nas.nasa.gov, or to Thorpe at thorpej@nas.nasa.gov.

Next Article | Contents | Main Menu | NAS Home



Recent advances in commercial disk drive technology have made off-the-shelf hardware an attractive solution for high-speed mass storage. This graph of the bandwidth shows the sustained transfer rate of the VWT disk array over an extended period using optimum block-size transfers. Notice that performance remains constant at around 300 megabytes per second (2.4 gigabits per second). The periodic downward spikes in the graph indicate lower bandwidth, and occur about every 660 blocks (or 4.3 gigabytes transferred). This lower bandwidth may be caused by periodic disk recalibration or by process scheduling in the workstation host, and remains a focus of ongoing investigation.

to the article

Computational Nanotechnology and Petaflops Computing Hold Promise for NASA Technology

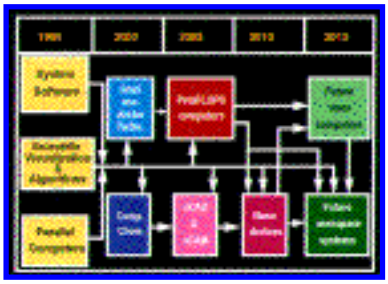
By Subhash Saini

At a recent press conference, NASA Administrator Dan Goldin encouraged Ames Research Center to take a lead role in promoting research and development of advanced, high-performance computer technology, including nanotechnology. NAS scientists, in collaboration with computational chemists at Ames, have begun work on nanotechnology problems -- and much more work lies ahead.

Molecular nanotechnology is an enabling technology of the 21st century. The concept is to place atoms in predetermined positions by manipulating them individually. Its principles were first espoused by Noble Prize winner Richard Feynman in 1959, who said: "The principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom."

Manufacturers of leading-edge microprocessors currently perform intensive large-scale simulations in the design and verification of semiconductor devices and microprocessors. Recently, the need for this simulation and modeling analysis has greatly increased, due in part to the ever-increasing complexity of these devices, as well as lessons learned from incidences such as the Intel Pentium chip divide error. Simulation, modeling, testing, and validation will be even more important for manufacturing molecular systems because of the complex specification of millions of atoms and thousands of assembly steps, as well as the simulation and modeling needed to ensure reliable, robust, and efficient fabrication of molecular devices.

Applications to NASA



Developing molecular nanotechnology will advance the exploration and settlement of space by overcoming current manufacturing technology limitations on reliability, performance, and affordability. For example, if structures of space vehicles are made of diamondoid-like materials instead of conventional materials, it is possible to make space vehicles that are perhaps 50 times stronger and 15 times lighter than those of today. Figure 1 shows a road map for designing future nanomechanical computers and molecular aerospace systems.

By embedding nanomechanical components in Earth-orbiting satellites, planetary probes, and piloted vehicles, the cost of future space programs could be reduced. It has been estimated that satellites using nanotechnology could measure 15 inches wide by 2 inches thick and weigh about 2 pounds. These satellites would require smaller and less expensive launchers. There is also the possibility of making ultra-light and ultra-strong solar sails.

Human exploration and development of space can probably become one of the foremost beneficiaries: nanorobots can be used for initial exploration as forerunners to human participation; small mission systems may exploit self-reproducing nanorobots; and the search for life can be carried out by using more efficient nanodevices and nanomachines.

Molecular Simulation Methods

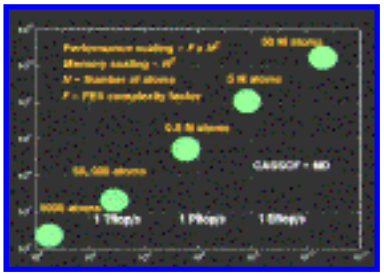
Molecular nanotechnology simulation requires approxi mately 200,000 to 100 million atoms. The most accurate methods available are necessarily limited in scope to smaller molecules. In such methods, CPU time is proportional to the fourth power of the basis functions, and to the seventh power of the number of valance electrons. Therefore, highly accurate computational chemistry ab initio calculations are almost impossible for molecular nanotechnology simulation.

To circumvent the computational expense, less accurate but faster methods must be used. Benchmark calculations have shown that the complete-active-space self consistent field (CASSCF) method, with MP2 correlation included, is quite accurate for 20-30 atoms. Potential Energy Surface (PES) is computed using CASSCF. The PES is converted to some analytical form, then used in molecular dynamics (MD) computations. Classical MD involves solving Hamilton's equations of motion for each atom in the system with respect to every other atom. The forces acting on the atom are the combination of forces due to bonds, van der Waals, and Coulomb forces. Calculating the Coulomb forces is typically the most computationally intensive because each atom interacts electrostatically with all other atoms in the system. Scientists at NAS are interested in a general class of MD simulation, which has two salient characteristics: short-range and long-range forces.

Short-range Forces

The first is short-range (local) forces; that is, each atom interacts only with other atoms that are within some cut-off distance. Many solids are modeled in this way because of electronic screening effects or because long-range interactions are truncated to avoid more costly computations. The local nature of the forces means that computational effort scales linearly with the number of atoms. There are several other approximate methods, such as the fast multipole expansion method and the cell multipole method.

Long-range Forces



The second is long-range (global) forces; that is, each atom interacts with every other atom. This is necessary when atoms undergo large displacements over the duration of a simulation. In a solid, reasons for including long-range forces could be diffusion, impurity, dislocation, or phonon-scattering interactions. This concept is important computationally because each atom's neighbors must be tracked as the simulation progresses, in order to take full advantage of the force locality. Two types of decompositions are commonly employed: atom and spatial. Assuming the above outlined strategy (ab initio + CASSCF + MD) NAS scientists have estimated the required computing power and disk space needed to perform long-range MD simulations for systems of interest to nanotechnologists. Results shown in Figure 2 underscore the need for petaflops computers in nanotechnology simulations.

Need Affordable Petaflops Computing

One of the biggest challenges in nanotechnology is the ability to simulate an entire assembly sequence of 100 million or more atoms into a "nano-object." This requires about a one-million- to one-billion-fold increase in today's computing power. There is a need for molecular nanotechnology computer-aided design and molecular nanotechnology computer-aided manufacturing systems. These systems are software environments for designing and manufacturing atomically accurate components comprised of 50,000 to 100 million atoms.

For nanotechnology to be successful, it will be essential to develop petaflops computing systems and associated software. Several NASA Grand Challenge problems, including nanotechnology simulations, have been identified that require petaflops (1 million billion floating point operations per second) and petabytes of storage space. To put this in perspective: the current high-end CRAY T90/16 has a peak performance of 28.8 gigaflops (1 billion floating point operations per second). Assuming that the price of a T90/16 is $30 million, it would take 34, 000 full systems at a cost of $1 trillion to achieve one petaflops of performance. In turn, the success of petaflops computing systems hinges on major advances in semiconductor technology. The obvious conclusion is that we need to develop affordable petaflops computer systems.

The NASA High Performance Computing and Communications Program is currently working towards achieving teraflops (Tflop/s -- 1 trillion floating-point operations per second) computing. There is consensus among academia, industry, and government that Tflop/s systems will not be adequate to solve many scientific and engineering problems that exist now, let alone for future applications. Achieving petaflops will require innovative technologies and possibly new computing paradigms. To advance this technology, a petaflops initiative is being pursued by NASA headquarters and involves the Department of Energy, the National Security Agency, the National Science Foundation, the Advanced Research Program Agency, and the Ballistic Missile Defense Organization. The current target for realization of a petaflops computer system is 2008.

NASA needs to analyze the system requirements of its future computational problems, such as

aerosciences, electromagnetics, Earth and space sciences, and molecular nanotechnology. These kinds of applications need to be studied to learn about issues such as degree of parallelism, memory size, data locality, interprocessor communication, and parallel I/O. The agency also needs to analyze future software requirements for these systems, including operating systems, parallel programming languages and tools, scientific subroutine libraries, performance tuning, I/O, and scientific visualization systems.

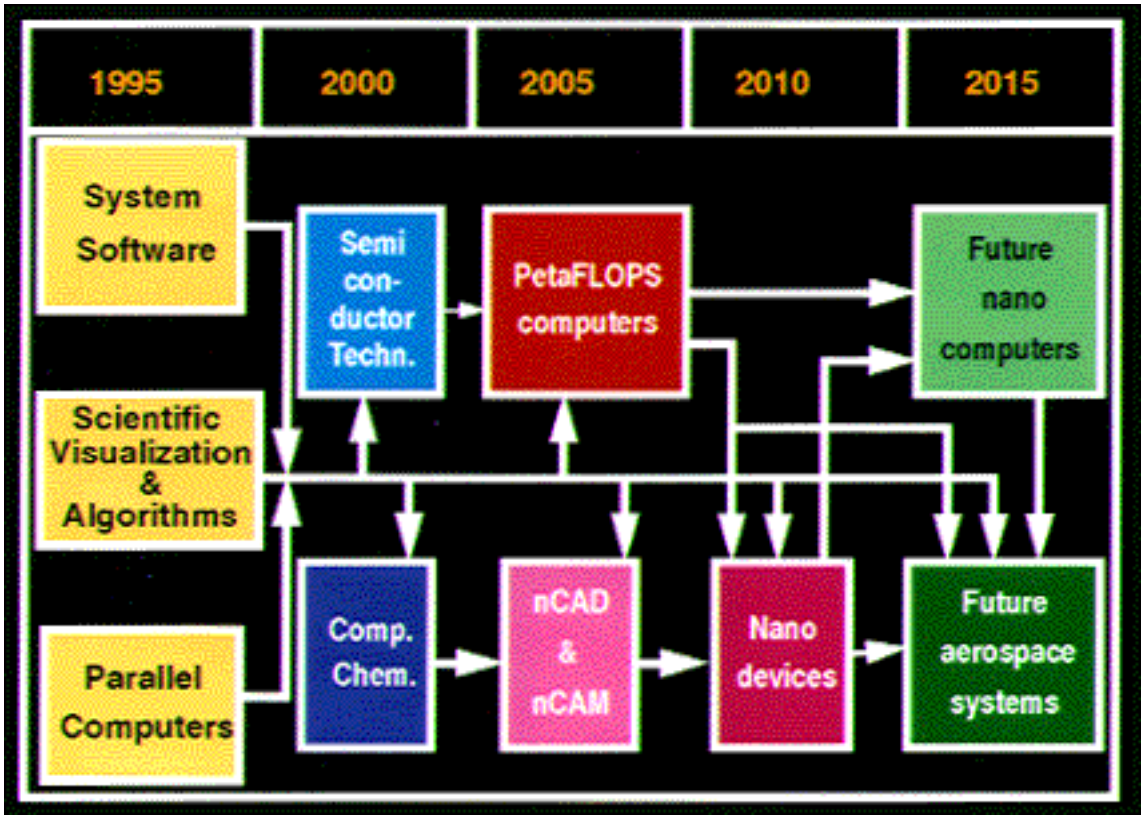
`Dramatic' Progress Required

Dramatic progress will be required in these areas to match the progress in computing speeds represented by these systems. Applications of interest to the scientific community that will most likely require petaflops computing include: multidisciplinary optimization of commercial aircraft designs; the direct numerical simulation of turbulence; computation of the radar scatter signature of aircraft bodies; full-scale simulation of the Earth's atmosphere and climate, and the operation of smart nanorobots for space exploration.

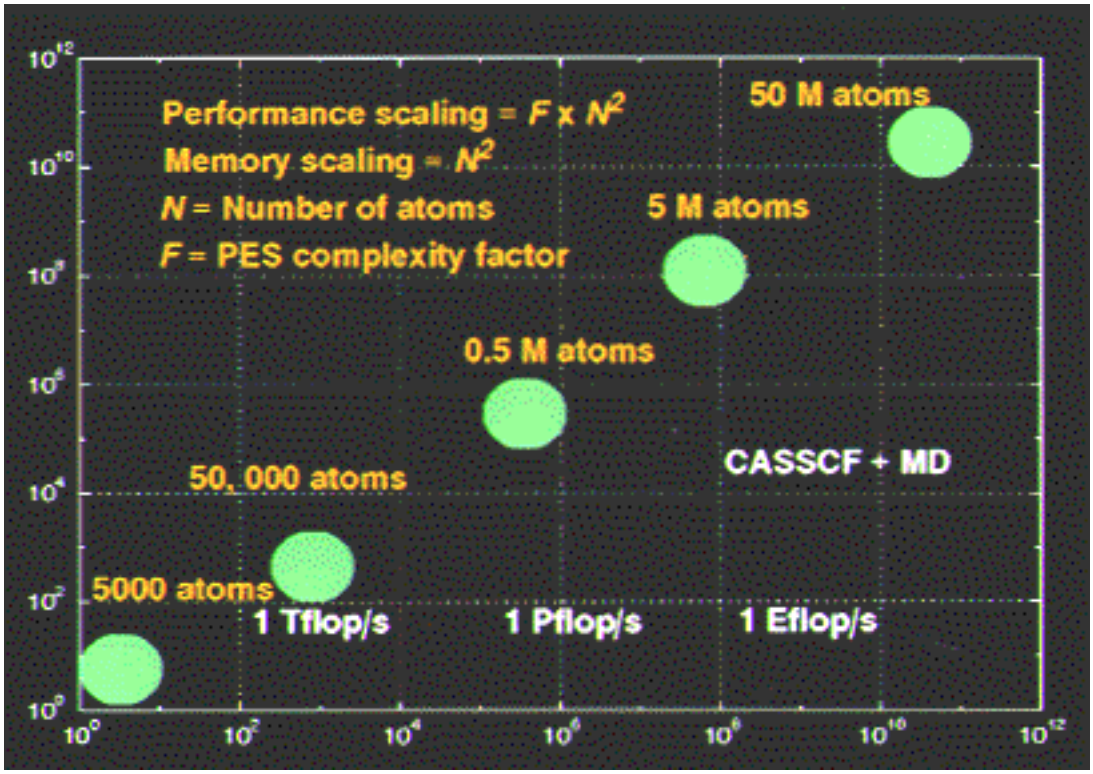
For details on computational nanotechnology work being done at NAS, send email to David Bailey at dbailey@nas.nasa.gov.

Subhash Saini is a scientist in the NAS Data Analysis Branch. He presented a paper, "Petaflops Computing and Computational Nanotechnology," at the Fourth Foresight Conference on Molecular Nanotechnology, held November 9-11, in Palo Alto, CA. In August he presented, "Prospects of Nanotechnology at NASA Ames Research Center." To request a video-tape of this presentation, send email to doc-center@nas.nasa.gov.

Next Article | Contents | Main Menu | NAS Home



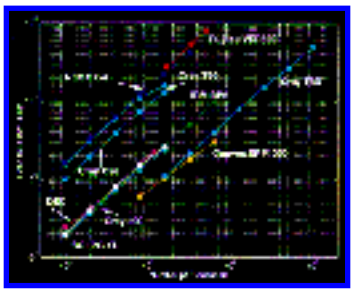
to the article



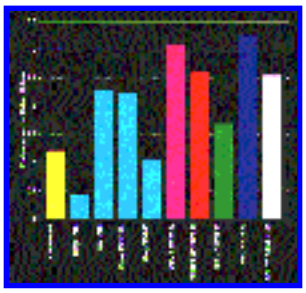
to the article

Latest NPB Results Published on Web

The latest NAS Parallel Benchmark (NPB) results, for the period April-October 1995, are now available on the World Wide Web. For detailed information, see the NAS Parallel Benchmarks home page.

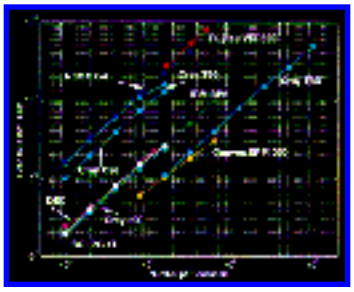


Eight benchmarks were used by vendors to measure performance for ten computer models. Figures 1 and 2 show the Block Triagonal (BT) benchmark results for scalability and price-performance for all models. In this benchmark, multiple independent systems of nondiagonally dominant block triagonal equations are solved with a 5x5 block size.

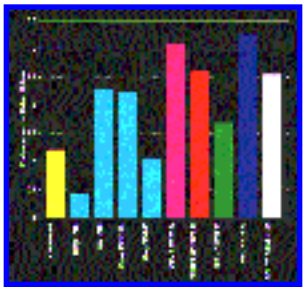


One of the best performing parallel computers, the NEC SX-4/32, was new to this round of NPB testing and incorporated CMOS [Complementary Metal Oxide Semiconductor] technology in its design. CMOS is "less expensive because it's air cooled," said Subhash Saini, a NAS scientist employed by MRJ Inc.

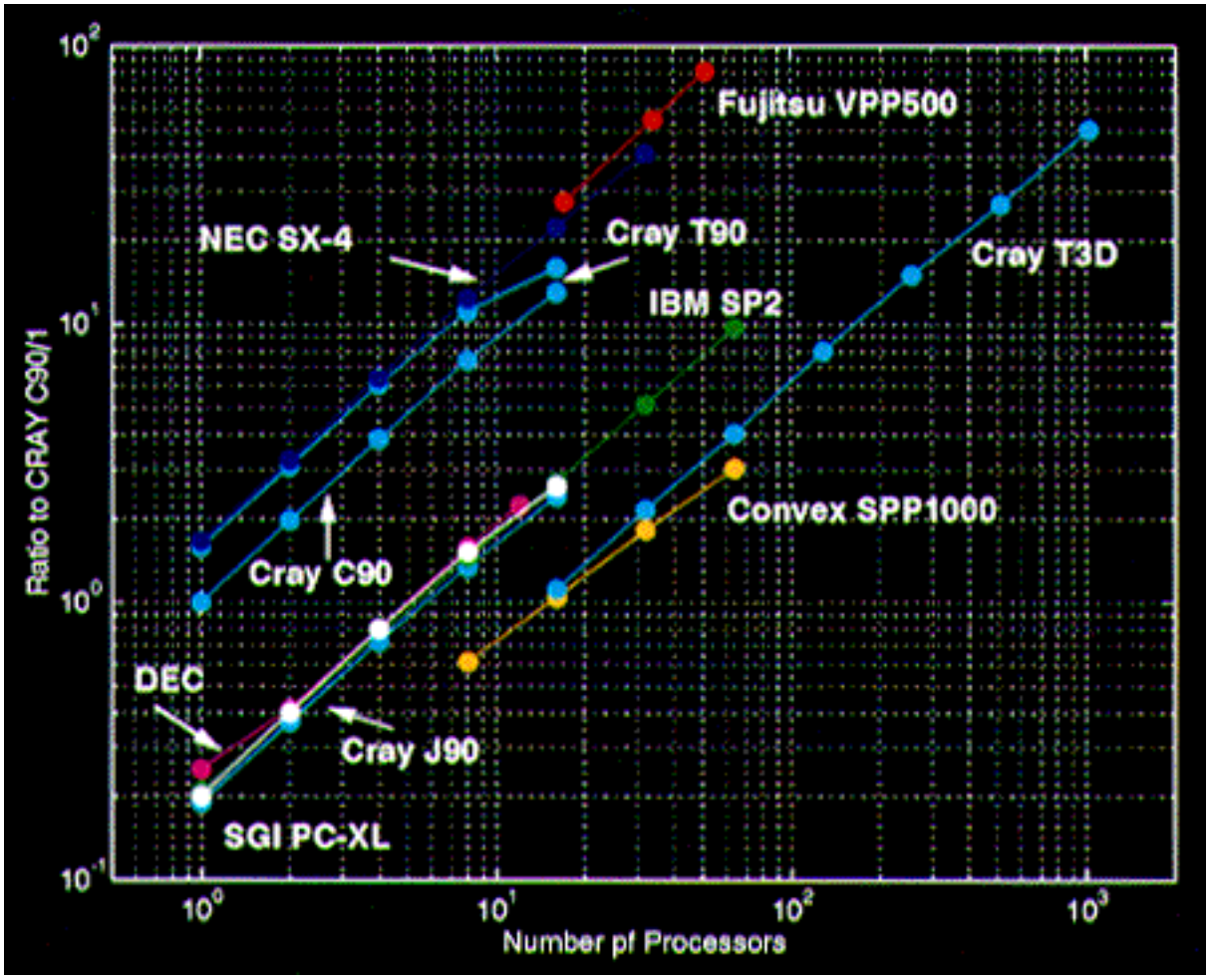
The following vendors submitted new models for testing: CRAY T916, IBM RS/6000 SP Wide-node2 (77 Mhz), and Silicon Graphics Inc. Power Challenge XL (90 Mhz). In addition, the CRAY J916 was measured for Class B benchmarks for the first time.

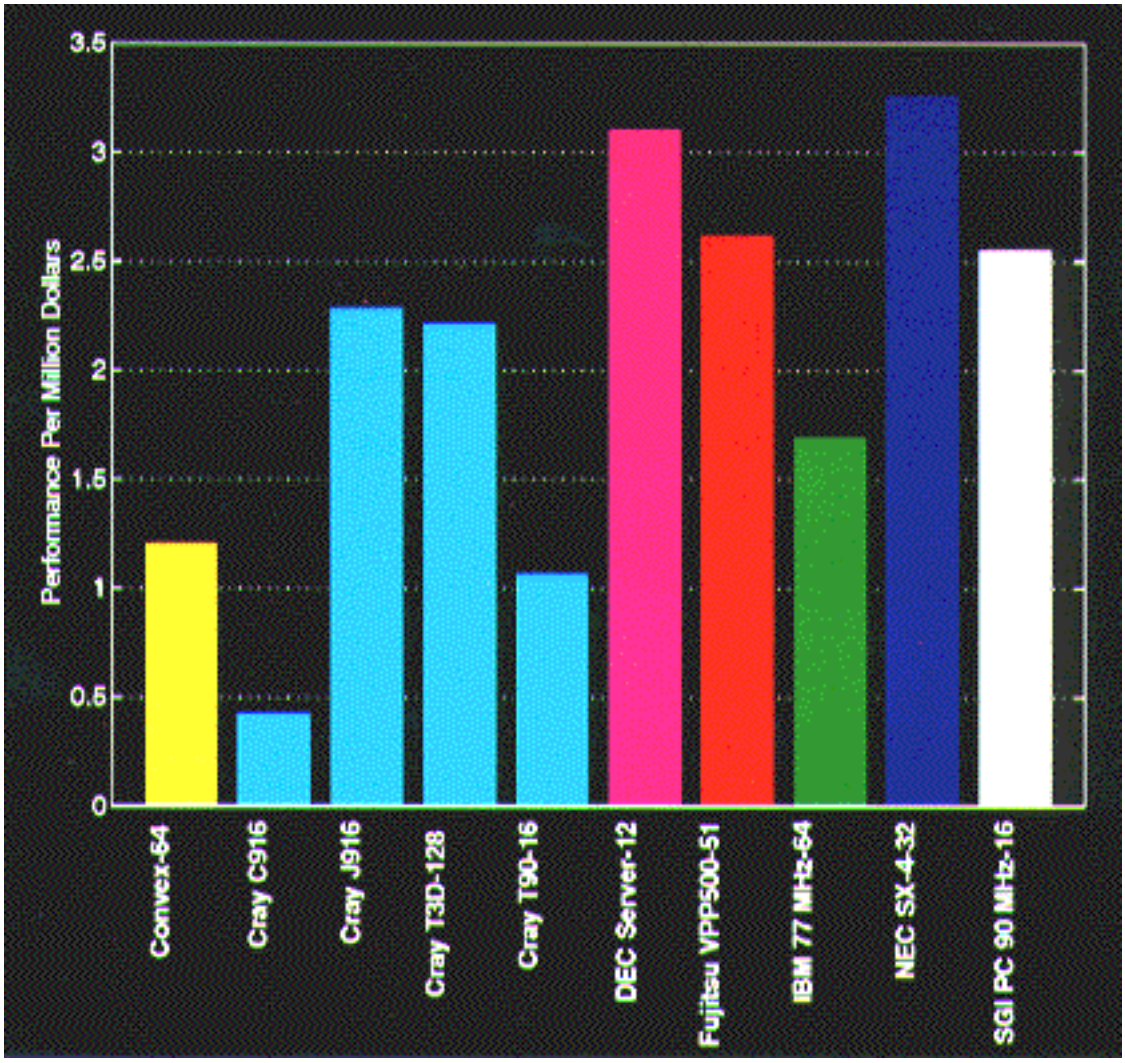


The Block Triagonal (BT) benchmark is used to show the scalability (Figure 1) and price-performance (Figure 2) of various high-performance computers. The Fujitsu VPP500, NEC SX-4/32, and IBM RS/6000 SP Wide-node2 achieved the best scalability results. The NEC SX-4/32, DEC Alpha Server-12, and Fujitsu VPP500-51 earned the best performance per million dollar investment. (Note that the price of the CRAY T3D does not include the cost of the front end.)



to the article


Fujitsu VPP500
NEC SX-4
Cray T90
Cray T3D
IBM SP2
Convex SPP1000
Cray C90
DEC
Cray J90
SGI PC-XL
Ratio to CRAY C90/1
Number pf Processors
10^2
10^1
10^0
10^-1
10^0
10^1
10^2
10^3


Performance Per Million Dollars
3.5
3
2.5
2
1.5
1
0.5
0
Convex-64
Cray C916
Cray J916
Cray T3D-128
Cray T90-16
DEC Server-12
Fujitsu VPP500-51
IBM 77 MHz-64
NEC SX-4-32
SGI PC 90 MHz-16

Serial Optimization Strategies on the IBM SP2

by Chuck Niggley and Mark Smith

This article is the first in a series that focuses on making the most of the IBM SP2's capabilities. Some of the methods for improving peak performance are peculiar to the NAS Facility's 160-node SP2 system and its companion 48-node system at NASA Langley Research Center.

Various serial optimization strategies have proven worthwhile for codes running on the SP2 POWER2 nodes. These strategies are summarized below:

- Use DOUBLE PRECISION whenever possible.
- Use the highest level of compiler optimization available that is still consistent with getting acceptably accurate results.
- Specify those options needed to tell the compiler to produce code for the POWER2 architecture of the nodes.
- Use numerical routines from libraries tailored specifically to the POWER2 architecture.
- Minimize I/O, especially if it involves NFS-mounted file systems, such as the home directories and /scratch disks.

Use DOUBLE PRECISION

The SP2's POWER2 nodes are double-precision (REAL*8 in Fortran, double in C) engines; internally, all floating-point computations are done in 64-bit precision. Any use of single-precision operands or results adds overhead in the form of format conversion. To get the best performance, either redeclare all computational operands as DOUBLE PRECISION, or use the -qautodbl command-line option for automatic conversion. In most cases, changing single-precision operands to double-precision floating point will decrease the compute time -- but it will also increase the code's storage requirements. If the amount of memory allocated per node is less than 100 megabytes (MB) after conversion, then this tradeoff will not be a problem.

Take Advantage of Optimization

The IBM XL family of compilers, which includes Fortran (mrf77, xlf90, or mpxlf), C (cc or mpcc), C++ (xlC or mpCC), and now HPF (xlhpf) uses a common set of optimization flags to give more detail about the POWER2 processor to the compiler optimization stage. No source statement reordering occurs at optimization level 2 (-O2). However, level 3 (-O3) combined with high-level optimization (-qhot) usually generates more efficient code. The only drawback is that level 3 may reorder arithmetic operations, which in turn may cause slight differences in rounding or truncation.

For the POWER2 nodes, the architecture-specific flags for the XL family of compilers are:

```
-qarch = pwr2   -qcache=d:level=1:size=256
```

Some nodes have more than 256 MB of memory; thus, an application addressing more than 256 MB of virtual memory may perform better on these nodes. The XL compilers generate code that addresses 256 MB for static (heap) virtual storage. To address a larger amount, use the following flag:

```
-bmaxdata:0xN0000000
```

The default is 256 MB, or N=0. If N=2, then a 512-MB address space is available. Similarly, if stack storage is too constraining (for example, your C or C++ program is running out of room for automatic variables), use the companion option:

```
-bmaxstack:0xM000000
```

Here's an example that summarizes the compiler optimization concepts: To compile a parallel code implemented with IBM Message Passing Library calls that require 512 MB of virtual address space, and with Fortran source code contained in the file bigprog.f, type the following:

```
mpxlf -o bigprog -O3 -qhot -qarch=pwr2\
  -qcache=d:level=1:size=256\
  -bmaxdata:0x20000000 bigprog.f
```

For more information, use the X Window-based InfoExplorer utility located on babbage1 or babbage2 (the SP2's front-end systems), or access online documentation for the individual compilers:

```
info -l xlf (or xlc or xlC)
```

A short man page for each compiler includes terse summaries of the compiler flags.

Use Tailored Numerical Libraries

One optimization strategy for implementing numerical algorithms is to make use of the available system libraries. This is particularly effective when the libraries have been expertly tuned for a specific system. For the SP2, the Engineering Scientific Subroutine Library (ESSL) contains a wide variety of solvers for both dense and sparse matrices, as well as FFTs. To get the tuned versions of these routines, specify -lesslp2 on the compiler command line. Similarly, the Optimization Subroutine Library (OSL) contains simplex and interior point methods for mathematical optimization.

Microsecond Timing with XL rtc( )

Before going into further detail using source code examples, let's review code timing on the SP2. Note that timings using TIME( ) or TIMEX( ) suffer from the low resolution of these UNIX timers. For microsecond resolution timing, use the rtc( ) command in IBM's XL compiler libraries instead of TIME or TIMEX. In XL/Fortran enter:

```
c  Fortran source code extract:
     REAL*8 starttime, READTIME, rtc
     starttime = rtc()
     READTIME  = starttime - rtc()
```

After executing the latter two statements, variable READTIME is set to the time it takes to call a subroutine on the SP2 node. Note that this timer is coarsely synchronized across the nodes by Network Time Protocol. (This topic will be covered in more detail in the January-February issue of NAS News.)

The following Fortran timing example compares two DAXPY operations implemented in Fortran versus the ESSL equivalent:

```
c  sample Fortran program extract:
c  optimize.f, source is in
      starttime = rtc()
      READTIME = starttime - rtc()
      initialization code not displayed
      starttime = rtc()
      do i=1, NS
             Y(I) = A(I)*x + b
             Z(I) = C(I)*x + b
      end do
      runtimef = rtc() - starttime - READTIME
      reinitialization code not displayed
      starttime = rtc()
      call daxpy(NS,x,A,1,Y,1)
      call daxpy(NS,x,C,1,Z,1)
      runtimel = rtc() - starttime - READTIME
```

```
        print *, ` runtime = ` , runtimef
        print *, ` library call = `, runtimel
Console output: TIMING RESULTS
   0:  runtime =  0.6526975631713867 19
   0:  library call =  0.1598317623138427 73
```

The above results show an almost three-fold increase in performance using the highly tuned ESSL code.

Minimize I/O, Maximize Local Disk

Since I/O operations are orders-of-magnitude slower than anything else that the code might be doing, the best optimization strategy is to avoid I/O as much as possible. On the SP2, your code has at least 80 MB of memory available on each node; so, try to keep the data needed in local memory. Once the data is there, do as many operations as possible before transferring more data to the system.

If you must do I/O, do it as quickly as possible by maximizing use of local disk space (/tmp) as well as PIOFS, the Parallel I/O File System. Each node has local disk space, which can be accessed as the /tmp file system; put as many files there as possible. Read or write from your home directory or any other NFS-mounted file system only as a last resort, since this is guaranteed to be unbearably slow, relative to the processor's time scale.

For more information on these optimization techniques, send email to scicon@nas.nasa.gov.

In October, the NAS parallel systems group was testing PIOFS, which was scheduled to be available in November. Look for details in the January-February issue.

Chuck Niggley leads the NAS scientific consulting group. Mark Smith is a member of IBM's technical staff working at the NAS Facility.

Next Article | Contents | Main Menu | NAS Home

Next Article | Contents | Main Menu | NAS Home

Nov-Dec 1995, Vol. 2, No. 13

Executive Editor: Marisa Chancellor

Editor: Jill Dunbar

Senior Writer: Elisabeth Wechsler

Contributing Writers: Walt Brooks, Eric Hibbard, Chuck Niggley, Subhash Saini, Mark Smith

Image Enhancements: Chris Gong, Steve Harding

Other Contributors: Bruce Blaylock, Nicholas Cardo, James Donald, John Hardman, Ed Hook, Terry Nelson, R.K. Owen, Alan Poston, Alan Powers, Marcia Redmond, Karl Schilke, Dani Thompson, Jason Thorpe

Editorial Board: Marisa Chancellor, Jill Dunbar, Chris Gong, Eric Hibbard, Kristina Miceli, Chuck Niggley, Serge Polevitzky, Pamela Walatka, Elisabeth Wechsler

NAS NEWS

Volume 2, Number 13 November – December 1996

NAS Ports Mass Storage to Lower-cost Platforms, Maintains Performance

by Elisabeth Wechsler

A project team at the NAS facility has developed a cheap and effective solution to mass storage, which traditionally has used specialized mainframe computers. "The goal is to port NASnet [NAS mass storage system] software to less expensive platforms and also to see if we can get the same kind of performance," said Alan Poston, staff scientist and project lead.

"By using multiple SCSI II [Small Computer System Interface] disk drives to build high-performance disk arrays, you gain the advantage of commodity pricing," he said. A disk array is a collection of disk drives administered by, for example, a number of SCSI buses.

"For about one-tenth of the current price, we get the same storage capacity at 100-150 percent performance improvement. Plus, we could avoid $1 million a year in maintenance costs for NAS's current mass storage system," said Ned Schilke, a student intern working on a part of the project involving the NAS Virtual Windtunnel (VWT).

Performance Exceeds Expectations

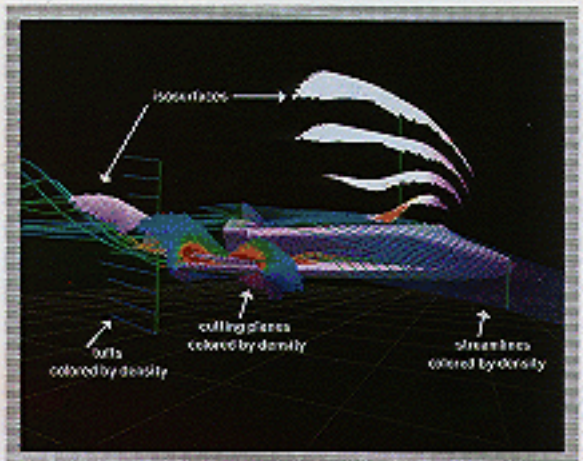
Performance levels are higher than expected. The original requirement was 300 megabytes per second (MB/sec) but 380 MB/sec was the performance level agreed to for the NAS VWT testbed. Initial data rates (at 270 MB/sec) exceeded those for Midrange Convex system (200 MB/sec), and subsequent results have exceeded 300 MB/sec (see accompanying graph on page 7), according to Schilke.

Using Silicon Graphics Inc.'s (SGI) IRIX 6.1 (a version of UNIX) with XFS, it's possible to set up drives in a striped configuration. "Striping is a divide-and-conquer technique," Schilke explained. "By spreading the data across many disks and reading it back in parallel, we can achieve very high aggregate bandwidth."

"The VWT uses the array as a buffer cache, reading CFD datasets at 300 MB/sec, and visualizing it in real time with special virtual reality software," he continued. "The VWT technique could be used in other applications that create large datasets for visualization in near real time."

Continued on page 7

THIS ISSUE



Visualization of the space shuttle during reentry, by the NAS Virtual Windtunnel (VWT). This image shows several visualization techniques: streamlines and tufts of the velocity vector field, as well as cutting planes and isosurfaces showing the density field. The visualizations are computed in real time by directly reading data from the CFD code as the data are flowing over a network. The VWT is one of many software projects that will be demonstrated in the NAS booth at Supercomputing '96, November 17-22, San Diego. See the NAS News "Special Issue" for more information on the SC'96 activities, also on the World Wide Web at: http://www.nas.nasa.gov/SC96. *Image by Steve Bryson.*

NAS Will Play Pivotal Role in Agency Supercomputing Consolidation, NCITE

by Walt Brooks
NAS Division Manager

NASA is at a strategic cusp in its evolution. Roles and missions within the agency are being clarified, and one important outcome is that Ames Research Center was recently named the lead center for Information Technology. To help clarify this new role, a plan has been submitted to NASA administrator Dan Goldin to initiate the NASA Center for Information Technology (NCITE), whose goal is the development of information technology capabilities necessary to "revolutionize NASA missions in the 21st century." As part of the effort to significantly streamline our operations and to consolidate roles and missions, Ames has also been named the lead center for NASA's scientific and engineering supercomputing.

The NAS division is currently well within Ames' role as the lead center for Information Technology. Within the last two years, the NAS Systems Division has taken on many new responsibilities, including setting agency technical and marketing major systems software development efforts for the High Performance Computing and Communications Program, as well as developing and operating the Numerical Aerodynamic Simulation Supercomputing Facility. This increase in scope and the resulting significant operational savings represent some initial steps toward realizing the NASA-wide supercomputing consolidation, now underway. *See related article on page 2.*

NAS will play a pivotal role in NCITE and other supercomputing efforts for the agency. The division will spearhead efforts in petaflops computing, and nanotechnology, and will continue to push for balanced systems that will enable us, at the turn of the century, to complete a full multidisciplinary analysis of an entire aircraft. Teraflops and petaflops performance will be necessary to implement desired resolution and accuracy. Although truly massively parallel processor architectures failed to live up to their promise, three prototypes fed performance, some underlying technological hurdles, and recent systems have—for limited applications—demonstrated sustained performance equal to or greater than the traditional vector symmetric multiprocessor model scale. To reach the teraflops and petaflops goals, an increase in parallel computing scalability will be required, by several orders of magnitude. The NAS Division will be at the heart of efforts to develop

Continued on page 2

Next Article | Contents | Main Menu | NAS Home

Visualization of the space shuttle dataset generated by the NAS Virtual Windtunnel (VWT). This image shows several visualization techniques: streamlines and tufts of the velocity vector field, as well as cutting planes and isosurfaces showing the density field. The visualizations are controlled in real time by moving various tools in the VWT. Two of the tools are shown here: rakes (vertical green lines) and sample points (green crosses). The VWT is one of many software products that will be demonstrated in the NAS booth at Supercomputing '95, December 3 -- 8, San Diego. See the NAS News Special Issue for more information on NAS's SC '95 activities; also see the World Wide Web at http://www.nas.nasa.gov/SC95.

Image by Steve Bryson.

Next Article | Contents | Main Menu | NAS Home